



Aalborg Universitet

AALBORG UNIVERSITY  
DENMARK

## Multiple Description Trellis-Coded Quantization of Sinusoidal Parameters

Larsen, Morten Holm; Christensen, Mads Græsbøll; Jensen, Søren Holdt

*Published in:*  
IEEE Transactions on Signal Processing

*DOI (link to publication from Publisher):*  
[10.1109/TSP.2008.928933](https://doi.org/10.1109/TSP.2008.928933)

*Publication date:*  
2008

*Document Version*  
Accepted author manuscript, peer reviewed version

[Link to publication from Aalborg University](#)

*Citation for published version (APA):*  
Larsen, M. H., Christensen, M. G., & Jensen, S. H. (2008). Multiple Description Trellis-Coded Quantization of Sinusoidal Parameters. *IEEE Transactions on Signal Processing*, 56(10), 5287 - 5291.  
<https://doi.org/10.1109/TSP.2008.928933>

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

### Take down policy

If you believe that this document breaches copyright please contact us at [vbn@aub.aau.dk](mailto:vbn@aub.aau.dk) providing details, and we will remove access to the work immediately and investigate your claim.

# Multiple Description Trellis-Coded Quantization of Sinusoidal Parameters

Morten Holm Larsen, *Student Member, IEEE*, Mads Græsbøll Christensen, *Member, IEEE*,  
Søren Holdt Jensen, *Senior Member, IEEE*

**Abstract**—A new scheme for sinusoidal audio coding named multiple description spherical trellis-coded quantization is proposed and analytic expressions for the point densities and expected distortion of the quantizers are derived based on a high-resolution assumption. The proposed quantizers are of variable dimensions meaning that any number of sinusoids can be quantized jointly for each audio segment whereby a lower distortion is achieved compared to previously published scalar spherical quantizers. The quantizers are designed to minimize a perceptual distortion measure subject to an entropy constraint for a given packet-loss probability. In experiments, the performance of the quantizers is assessed and compared to the corresponding single description spherical quantizer and associated bounds under various conditions and is found to increase robustness towards packet-loss.

## I. INTRODUCTION

In recent years, there has been a significant interest in parametric audio coding in both academia and standardization bodies. Parametric audio coding is based on the notion that most audio signals can be efficiently described by a few physically or perceptually meaningful parameters. This process can be seen as vector quantization using a highly structured codebook that allows for computationally efficient implementation. Perhaps the most common incarnation is sinusoidal coding where the individual audio segments are modeled as sums of sinusoids each of which being characterized by an amplitude, a phase and a frequency combining to form a point in a spherical coordinate system. For each segment of audio, the task is to find the parameters best describing the segment and to quantize these parameters whereby transmission over channels of limited capacity is facilitated. Various computationally efficient ways of finding the parameters minimizing a perceptual distortion measure exist (see, e.g., [1]) and the question of optimal quantization of these parameters has been addressed recently. The so-called polar and spherical quantizers of [2]–[4] have proven successful in terms of achieved quality and computational complexity by quantization of the parameters of each sinusoid independently. The quantizers were designed to minimize a perceptual distortion measure subject to an entropy constraint based on a high-resolution assumption, i.e., a high number of bits per sinusoid whereby analytic expressions for the point densities of the quantizers were derived. In [5], the spherical quantizers of [4] were improved by joint quantization of the parameters of a variable number of sinusoids. Under a high-resolution assumption, the optimal point densities of the proposed quantization scheme, named spherical trellis-coded quantization (STCQ), were derived for a given entropy.

M. H. Larsen, M. G. Christensen and S. H. Jensen are with the Department of Electronic Systems, Aalborg University, Denmark. (phone: +45 9635 8684, fax: +45 9635 1583 and email: {mhl,mgc,shj}@es.aau.dk)

M. G. Christensen is supported by the Parametric Audio Processing project, Danish Research Council for Technology and Production Sciences grant no. 274-06-0521.

In services such as speech coding or audio streaming over unreliable networks like the Internet, the transmitted audio parameters should be protected to compensate for packet-losses. One method intending to do this is multiple description coding [6] where several complementary coarse descriptions of the audio signal are constructed and transmitted. This way, graceful degradation is achieved when packets are lost while high quality reconstruction is possible when all packets are received. Multiple description coding has recently been applied to audio in the form of transform coding [7], [8] and low-delay coding using pre- and post-filtering [9]. The multiple description coding schemes of [8] and [9] are based on the multiple description lattice vector quantization (MDLVQ) of [10] and [11], respectively. The MDLVQ is a very computational efficient quantizer and flexible in the sense that it can be designed analytically, except for the index assignment. A limitation of the MDLVQ is that the dimension of the vector must be *fixed* and known *a priori*. Therefore, the MDLVQ cannot readily be applied to the problem of joint quantization of sinusoidal parameters. On the other hand, multiple description trellis-coded quantizer (MDTCQ) of [12]–[14] can handle variable dimensions but requires training for a particular combination of entropy constraint and packet-loss probability using the Lloyd algorithm.

In this correspondence, we extend the spherical quantizers of [4] to multiple descriptions to obtain robustness towards packet-losses. Furthermore, we also propose joint quantization of sinusoidal parameters using trellis-coded quantization. The proposed quantization scheme, called multiple description spherical trellis-coded quantization (MDSTCQ), is based on high-resolution theory, from which analytic expressions for the expected distortion and point densities are derived for a given target entropy and packet-loss probability. The MDSTCQ is based on a new quantization scheme named modified multiple description trellis-coded quantization (MMDTCQ) that can be analytically designed from its point density given a packet-loss probability. Interestingly, the proposed MMDTCQ scheme can readily be applied to a larger class of problems than considered here. Throughout the deviation of the MDSTCQ, some assumptions and approximations are applied with respect to both the source and the multiple description coding part. Most of these have already been discussed at great length in prior works [2]–[4] and we will therefore, for the most part, refrain from any further discussion of this.

The rest of this correspondence is structured as follows: In Section II, we introduce the definitions, fundamentals and the objective of multiple description audio coding. Then, in Section III, the proposed MMDTCQ is presented. The optimal point densities are derived for the MDSTCQ in Section IV, and in Section V the experimental results are presented. Finally, we conclude on our work in Section VI.

## II. FUNDAMENTALS

We start this section by introducing the mathematical problem statement of quantization in parametric audio coding based on a perceptually relevant distortion measure. Let the audio signal  $x$  at sample time  $n$  be represented as  $x(n) \approx \sum_{l=1}^L a_l \sin(\nu_l n + \phi_l)$ , where  $L$  is the number of sinusoidal components and  $a_l, \phi_l, \nu_l$  are the amplitude, phase and frequency of the  $l$ 'th component, respectively, with  $a_l \geq 0$  and  $\phi_l, \nu_l \in [0, 2\pi)$ . The quantization distortion consists of the contributions from the individual components and the cross-terms between the components. Assuming a sufficiently large window length  $W$ , so the sinusoids become orthogonal, the total expected distortion can be approximated as the sum over the  $L$  expected distortions for the individual components, denoted  $E[D]$ , with  $E[\cdot]$  being the expectation operator. Therefore, we will in the rest of this paper be concerned with the quantization of a single set of parameters  $(a, \phi, \nu)$  thus ignoring the subscript  $l$ . The present work is based on a perceptual distortion measure [15] successfully applied to audio coding (see, e.g., [4], [16]–[18]). The distortion measure is defined as

$$D = \frac{1}{2\pi} \int_0^{2\pi} \mu_{x(a,\phi,\nu)}(\omega) |E(\omega)|^2 d\omega, \quad (1)$$

with  $E(\omega)$  denoting the discrete-time Fourier transform of the windowed error, i.e.,  $E(\omega) = \sum_{n=n_0}^{n_0+W-1} w(n)(x(n) - \tilde{x}(n))e^{-j\omega n}$  where  $w(n)$  is the window,  $\mu_{x(a,\phi,\nu)}(\omega)$  is the perceptual weighting function calculated from the audio signal  $x$  parameterized by  $(a, \phi, \nu)$ . Furthermore,  $\tilde{x}$  is the reconstructed audio signal based on the quantized parameters  $(\tilde{a}, \tilde{\phi}, \tilde{\nu})$ . For more information on sinusoidal parameter estimation based on the perceptual distortion measure in (1) see [1] and the references therein. Next, we introduce the quantization errors  $\epsilon_a = a - \tilde{a}$ ,  $\epsilon_\phi = \phi - \tilde{\phi}$ , and  $\epsilon_\nu = \nu - \tilde{\nu}$ . Then, assuming high-resolution and a smooth masking curve combined with the prior assumption that  $W$  is large, the perceptual distortion can be approximated as (see, e.g., [4])

$$D \approx \frac{\mu_{x(a,\phi,\nu)}}{2} \left( \|w\|^2 (a^2 + \tilde{a}^2) - 2a\tilde{a} \times \sum_{n=n_0}^{n_0+W-1} w^2(n) \cos(\epsilon_\nu n + \epsilon_\phi) \right), \quad (2)$$

where  $\|w\|^2 = \sum_{n=n_0}^{n_0+W-1} w^2(n)$ . Here, we have assumed that  $W$  is large whereby the perceptual weighting function reduces to a scaling  $\mu_{x(a,\phi,\nu)}$ . Similarly to [2]–[4], we assume the perceptual weighting function to be quantized and transmitted as side information. To the best of our knowledge, the problem of joint quantization of the perceptual weighting function and the sinusoidal parameters remains unsolved and we will defer from any further discussion of this difficult problem in this correspondence.

By assuming that the phase is defined in the middle of the segment, i.e.,  $n_0 = -\frac{W}{2}$ , and by applying a truncated Taylor expansion around zero,  $\cos(y) \approx 1 - y^2/2$ , the distortion can be shown to be

$$D \approx \frac{\mu_{x(a,\phi,\nu)}}{2\|w\|^{-2}} \left( \epsilon_a^2 + a\tilde{a} (\epsilon_\phi^2 + \epsilon_\nu^2 \sigma^2) \right), \quad (3)$$

with  $\sigma^2 = \frac{1}{\|w\|^2} \sum_{n=-W/2}^{W/2-1} w^2(n)n^2$ . We observe from (3) that the amplitude, phase and frequency can be quantized independently using the  $l_2$ -norm by assuming a high-resolution, i.e.,  $a\tilde{a} \approx a^2$ . This provides inspiration to the proposed multiple description spherical trellis-coded quantization (MDSTCQ)

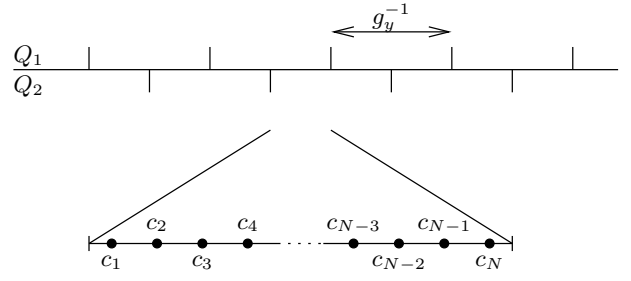


Fig. 1. Illustration of the structure of the MMDTCQ. The upper vertical lines illustrate the Voronoi regions of the two side quantizers  $Q_s$  while the lower dots illustrate the refined reconstruction points.

scheme consisting of three multiple description quantizers, one for each of the parameters  $a$ ,  $\phi$  and  $\nu$ . Multiple description is a graceful degradation scheme where a low reconstruction quality is obtained when only a few descriptions are received and a high reconstruction quality is obtained receiving many descriptions. We here focus on the common case of two descriptions. In this case, a low quality description is obtained when only one of the descriptions is received and the resulting distortion is called the side distortion and is denoted as  $D_s$ , with  $s = \{1, 2\}$ . The low quality reconstructions of  $(a, \phi, \nu)$  are written as  $(\tilde{a}_s, \tilde{\phi}_s, \tilde{\nu}_s)$ . When both descriptions are received, the resulting distortion is referred to as the central distortion  $D_c$  and is based on the reconstruction point  $(\tilde{a}_c, \tilde{\phi}_c, \tilde{\nu}_c)$ . The aim of this work is to protect the sinusoidal parameters transmitted over a packet erasure channel where packets are dropped independently with probability  $p$ . The average distortion in such a network is given by

$$E[D] = (1-p)^2 E[D_c] + 2p(1-p) E[D_s] + p^2 E[x^2], \quad (4)$$

where  $E[x^2]$  is the variance of the audio signal and balanced side distortion  $E[D_1] = E[D_2]$ . We will derive the optimal design of the proposed MDSTCQ scheme, for a given packet-loss probability, such that the perceptual distortion is minimized subject to an entropy constraint.

## III. MODIFIED MULTIPLE DESCRIPTION TRELLIS-CODED QUANTIZATION

Inspired by the structure of the MMDSQ in [19], we propose a modified multiple description trellis-coded quantizer (MMDTCQ) composed of two stages. The first stage is to quantize the input signal  $y$  using two uniform side quantizers,  $Q_1$  and  $Q_2$  which are offset to each other and have step-sizes equal to the reciprocal value of the point density  $g_y$ . This is depicted in Fig. 1. In the second stage, we perform joint quantization using the joint Voronoi region, which is half the size of the Voronoi region of the side quantizers, similarly to the MMDSQ of [19]. Here,  $N$  reconstruction points are sorted as  $c_1 < c_2 < \dots < c_N$  and partitioned into four subsets  $C_q = \{c_q, c_{q+4}, \dots, c_{N-4+q}\}$ , where  $q = \{1, 2, 3, 4\}$ . In trellis-coded quantization, the transitions in the trellis specify  $C_q$  by one bit and  $\log_2(N/4)$  bits are used to specify the quantization index within  $C_q$ . Since this refined trellis information is only utilized when both descriptions are received, it is easily split into the two descriptions whereby the otherwise difficult index assignment (see, e.g., [10], [11], [20]) is solved in a straightforward manner. For more information about trellis-coded quantization we refer the interested reader to e.g., [21]. Assuming high-resolution such that the probability density function (pdf)

inside the Voronoi regions  $V_i$  of the side quantizers  $Q_s$  is uniform, the mean square-error distortion for each side quantizer can be determined by summing over all Voronoi regions,  $i \in I$ ,

$$E[D_s] = \sum_{i \in I} \int_{V_i} f_Y(y) \epsilon_y^2 dy \approx \frac{g_y^{-2}}{12}, \quad (5)$$

where  $f_Y(y)$  is the source pdf and  $\epsilon_y = y - \tilde{y}$ . In [19] it was shown that the side distortion is reduced by a factor of  $(2N)^2$ , but unfortunately, an exact expression of the expected distortion does not exist for the trellis. Therefore, we will employ the approximation proposed in [21] where the expected distortion is written as the distortion for a uniform quantizer corrected by a factor  $\Gamma$ . This factor depends on the trellis structure, dimension and the number of states and is roughly independent of the source pdf and encoding rate. Also, assuming high-resolution, the central distortion can be written as

$$E[D_0] \approx \frac{\Gamma}{(2N)^2} E[D_s]. \quad (6)$$

The  $\Gamma$  factor has been numerically determined by simulations in [5].

Under high-resolution assumption and assuming a half bit per description to specify the trellis transition,  $C_q$ , the entropy of a MMDTCQ per description can be approximated by  $H_s \approx h(Y) + \int f_Y(y) \log_2(g_y) dy + \frac{1}{2} \log_2\left(\frac{N_y}{2}\right)$ , where  $h(Y)$  is the differential entropy of the source  $Y$ . Note that we here denote random variables by upper case letters whereas realizations are denoted by lower case. The complexity of the MMDTCQ can be quantified as follows: In the first stage, two uniform quantizations are performed or, alternatively, one quantization on the subdivided Voronoi region and, subsequently, one check on whether the index number is even or odd. The complexity of the second stage is similar to four uniform quantizers plus  $2S + 4$  additions, 4 multiplications and  $S$  comparison of two real valued numbers with  $S$  being the number of states.

#### IV. MULTIPLE DESCRIPTION SPHERICAL TRELLIS-CODED QUANTIZATION

We start this section by introducing the details of the proposed MDSTCQ coding scheme consisting of three MMDTCQs, one for each of  $a$ ,  $\phi$  and  $\nu$ . In deriving the optimal MDSTCQ design, we need expressions for the quantization point densities and the number  $N$  for the three MMDTCQs. To obtain these, we first introduce the joint pdf  $f(A, \phi, \nu)$  and let  $V_i$  and  $I$  denote the Voronoi regions and index sets for the respective quantizers as indicated by subscripts  $a, \phi, \nu$ . We can now express the expected side distortion as

$$\begin{aligned} E[D_s] &\approx \sum_{i_a \in I_A} \sum_{i_\phi \in I_\phi} \sum_{i_\nu \in I_\nu} \int_{V_{i_a}} \int_{V_{i_\phi}} \int_{V_{i_\nu}} f(A, \phi, \nu) \frac{1}{2} \mu_x(\tilde{\nu}) \\ &\|w\|^2 (\epsilon_a^2 + a\tilde{a}_s (\epsilon_\phi^2 + \epsilon_\nu^2 \sigma^2)) dad\phi d\nu \approx \frac{\|w\|^2}{24} \int_A \int_\phi \int_\nu \\ &f(A, \phi, \nu) \mu_x(\tilde{\nu}) (g_a^{-2} + \tilde{a}_s^2 (g_\phi^{-2} + \sigma^2 g_\nu^{-2})) dad\phi d\nu \end{aligned} \quad (7)$$

by assuming that  $\mu_{x(a, \phi, \nu)}$  is constant over the joint Voronoi region of  $a, \phi, \nu$ . We have assumed high-resolution such that  $a\tilde{a}_s \approx \tilde{a}_s^2$  and the probability mass function of the reconstruction points,  $\Pr(\tilde{a}_{i_a}, \tilde{\phi}_{i_\phi}, \tilde{\nu}_{i_\nu})$  can be found from the joint pdf as  $f(\tilde{a}_{i_a}, \tilde{\phi}_{i_\phi}, \tilde{\nu}_{i_\nu}) g_a^{-1} g_\phi^{-1} g_\nu^{-1}$  (see [4] for more details on this). Here, the quantization point densities for  $a, \phi, \nu$  are written as  $g_{\{a, \phi, \nu\}}$ , though they at this point still depend on  $a, \phi, \nu$ .

Similarly, we can express the expected central distortion as

$$\begin{aligned} E[D_0] &\approx \frac{\|w\|^2}{96} \Gamma \int_A \int_\phi \int_\nu f(A, \phi, \nu) \mu_x(\tilde{\nu}) \left( \frac{g_a^{-2}}{N_a^2} \right. \\ &\left. + \tilde{a}_0^2 \left( \frac{g_\phi^{-2}}{N_\phi^2} + \sigma^2 \frac{g_\nu^{-2}}{N_\nu^2} \right) \right) dad\phi d\nu, \end{aligned} \quad (8)$$

where, for simplicity, we have assumed an equal trellis structure,  $\Gamma = \Gamma_{\{a, \phi, \nu\}}$ . Also,  $N_a, N_\phi, N_\nu$  are the number of reconstruction points for the various refined quantizers. Next, assuming high-resolution we can write the entropy for each description as

$$\begin{aligned} H_s &\approx h(A, \Phi, \Upsilon) - \frac{3}{2} + \iiint f(a, \phi, \nu) (\log_2(g_a g_\phi g_\nu) \\ &+ \frac{1}{2} \log_2(N_a N_\phi N_\nu)) dad\phi d\nu, \end{aligned} \quad (9)$$

with  $h(A, \Phi, \Upsilon)$  being the differential entropy. To simplify the notation, we introduce  $\tilde{H}_s = H_s - h(A, \Phi, \Upsilon)$  and write the cost function as  $J = (1-p)^2 E[D_0] + 2p(1-p) E[D_s] + \lambda \tilde{H}$  with  $\lambda$  being the Lagrange multiplier.

From (7) and (8), we observe that there is the somewhat subtle problem that the side and central distortions depend on the reconstructed values  $\tilde{a}_s$  and  $\tilde{a}_0$ . Ideally, the point densities also depend on the amplitude  $a$  and in [4] it was argued that the amplitude can be replaced by its reconstruction due to the high-resolution assumption whereby a feasible solution is obtained. Similarly, we here use  $\tilde{a}_s$  in lieu of  $\tilde{a}_0$  (later we will evaluate the loss, if any, of doing this). We now proceed to minimize this cost function by taking the derivative with respect to  $g_a, g_\phi, g_\nu, N_a, N_\phi$  and  $N_\nu$ . This results in  $N_a = N_\phi = N_\nu = \sqrt{\Gamma(1-p)/8p} \triangleq N$  and the following expressions for the point densities:

$$g_a^2 = \frac{p(1-p) \mu_x(\tilde{\nu}) \|w\|^2}{12\lambda \log_2(e)} \quad (10)$$

$$g_\phi^2 = g_a^2 \tilde{a}_s \quad (11)$$

$$g_\nu^2 = g_a^2 \tilde{a}_s \sigma^2. \quad (12)$$

We note in passing that  $N$  is independent of the source differential entropy as well as the target entropy, like the number of refined lattice points in MDLVQ (see, e.g., [10]). Returning now to our derivation, we insert the last three equations and the expression for  $N$  into the definition of  $\tilde{H}_s$ , whereby we obtain the optimal  $\lambda^*$  as

$$\lambda^* = \frac{Np(1-p) \|w\|^2}{12 \log_2(e)} 2^{\frac{2}{3}(h(a, \phi, \nu) - H_s + \log_2(\sigma) + 2\varrho) + 1 + \psi}, \quad (13)$$

where  $\varrho = \int f_A(a) \log_2(a) da$  with  $f_A(a)$  being the pdf of  $a$  and  $\psi = \iiint f(a, \phi, \nu) \log_2 \mu_{x(a, \phi, \nu)} dad\phi d\nu$ . Substituting  $\lambda^*$  into the expressions of  $g_a, g_\phi$  and  $g_\nu$ , we get the optimal point densities:

$$g_a = \left( \frac{\mu_x(\tilde{\nu})}{2N} \right)^{\frac{1}{2}} 2^{\frac{1}{3}(H_s - h(a, \phi, \nu) - \log_2(\sigma) - 2\varrho) - \frac{\psi}{2}} \quad (14)$$

$$g_\phi = \tilde{a}_s g_a \quad (15)$$

$$g_\nu = \sigma \tilde{a}_s g_a. \quad (16)$$

Next, by inserting (14)-(16) into (7), (8) and (4), we can finally determine the expected distortion as

$$\begin{aligned} E[D] &\approx \|w\|^2 \sqrt{\frac{\Gamma p}{8}} (1-p)^{\frac{3}{2}} 2^{\frac{2}{3}(h(a, \phi, \nu) - H_s + \log_2(\sigma) + 2\varrho) + \psi} \\ &+ p^2 E[x^2]. \end{aligned} \quad (17)$$

The point densities for  $\phi$  and  $\nu$  can be seen to be functions of  $\tilde{a}_s$  (or  $\tilde{a}_0$ ), i.e., the point densities depend on the reconstructed amplitude. Since the fundamental notion of multiple



description coding is to transmit complementary descriptions that combined lead to a reduced distortion, the reconstructed amplitudes in each of the descriptions will be different, i.e.,  $\tilde{a}_1 \neq \tilde{a}_2$ , leading to phase and frequency quantizers having slightly different resolutions for each description. To arrive at a feasible scheme, we encode  $\phi_s$  and  $\nu_s$  based on  $\tilde{a}_s$  for each of the two descriptions with  $s = \{1, 2\}$ . In the ideal case, the point densities for each description are equivalent and the quantizers can be perfectly offset as illustrated in Fig. 1, in which case the reconstruction points of the joint description can be obtained as the mean of the reconstruction points of the two descriptions plus the contribution from the refined trellis-coded quantization, e.g.,  $\tilde{\phi}_0 = \frac{\tilde{\phi}_1 + \tilde{\phi}_2}{2} + \tilde{\phi}_{TCQ}$ . In our case, however, the resolution of the two quantizers are slightly different, and we propose to deal with this in the following way: We consider the two descriptions as random variables from which we seek to estimate the mean. Due to the slightly different amplitudes  $\tilde{a}_1$  and  $\tilde{a}_2$ , the two observations have been subjected to additive noise having different variances. Therefore, the mean can be obtained for the phase as

$$\tilde{\phi}_0 = \tilde{\phi}_1 \zeta_\phi + \tilde{\phi}_2 (1 - \zeta_\phi) + \tilde{\phi}_{TCQ}, \quad (18)$$

where  $0 \leq \zeta_\phi \leq 1$  is a weight. Since the quantizers are uniform, the observation noise can be modeled as uniform random variables having a variance corresponding to their Voronoi regions and the optimal weight  $\zeta_\phi$  can then easily be determined from the two point densities as  $\zeta_\phi = g_{\phi,1}^2 / (g_{\phi,1}^2 + g_{\phi,2}^2)$  and similarly for the frequency  $\nu$ .

## V. EXPERIMENTAL RESULTS

We will now proceed to evaluate the performance of the MMDTCQ before we evaluate the proposed MDSTCQ. Based on the method described in Section III, we obtain the empirical performance curve of the MMDTCQ and compare it to the MMDSQ [19] and the corresponding theoretical bound [22]. We draw 10,000 realizations from a Gaussian distributed source, which are then quantized by a MMDTCQ and a MMDSQ. Both designs have an entropy below three bits per description and the MMDTCQ is a 256 state trellis with dimension 1,000. The performances of the MMDTCQ, MMDSQ and the optimal theoretically attainable (OPTA) bound are depicted in Fig. 2. It can be seen that the MMDTCQ is outperforming the MMDSQ and that the gap between the practical and the theoretical bound is narrowing. Note that for low side distortion where  $N$  is small, the two schemes have not been evaluated, since the MMDTCQ requires  $N \geq 4$  and MMDSQ requires  $N \geq 2$ .

The three following experiments are based on synthetic audio, generated using a statistical model similar to that employed in [4]. Specifically, the amplitudes and frequencies are generated from a Rayleigh pdf, i.e.,  $f_Y(y) = \beta^{-2} y e^{-(y^2 \beta^{-2}/2)}$ , with  $\beta = \{1000, 0.25\}$ . The phase, on the other hand, is uniformly distributed in the interval  $[0, 2\pi)$ . Note that the amplitudes, phases and frequencies are statistically independent. In this work we focus on the performance gain achieved by joint quantization and we will therefore, for simplicity, set the perceptual weighting function to one and use a rectangular window where  $W = 1023$ . We remark that the trellis is initialized in zero-state and that no side information is needed.

We will now investigate the impact of the number of jointly quantized sinusoids and the number of states in the trellis on the expected distortion. Furthermore, we will illustrate the impact of the estimation of  $\phi_0$  and  $\nu_0$  at the decoder as discussed

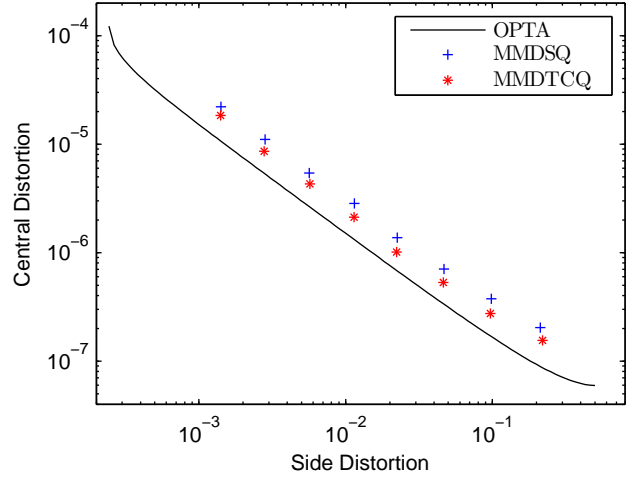


Fig. 2. MMDTCQ (star) and MMDSQ (plus) performance for a memoryless Gaussian source and 3 bits entropy per description. Also shown is the theoretically bound (solid).

in Section IV. We generate 100,000 triplets  $(a, \phi, \nu)$ , set the target entropy to 15 bit/sinusoid per description, using 256 state trellis with  $N = 4$  and quantize the triplets as described in Section IV. As explained, the optimal reconstruction of MDSTCQ is not feasible, but a feasible solution can be obtained by estimating  $\tilde{\phi}_0$  and  $\tilde{\nu}_0$ . The performance of two non-feasible and the single feasible MDSTCQ schemes are shown on Fig. 3 for a range of dimension. From Fig. 3 it can be seen that we gain about 1.2 dB when increasing the dimension from 10 to 10,000. Furthermore, it can be seen that there is a 2.2 dB gap between the non-feasible method and the feasible MDSTCQ method. Further simulations have shown the gap to be source and rate dependent. But it is important to note that this gap is due to a non-feasible assumption in the derivation of the expected distortion and not necessarily a suboptimality in the practical scheme. How to incorporate the feasibility criterion in the derivation is subject of future research.

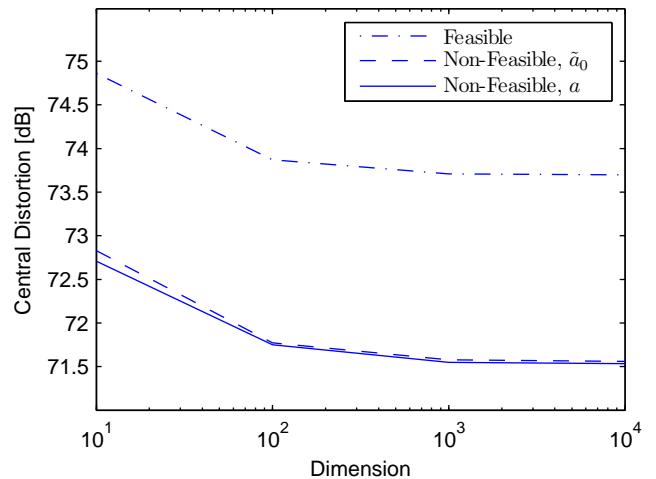


Fig. 3. Performance of the feasible MDSTCQ scheme (dash-dotted) compared to those of the two non-feasible MDSTCQ schemes where  $a$  is known in both side decoders (solid) and where  $\tilde{a}_0$  is known in the two decoders (dashed).

In the next experiment, we compare the performance of the MDSTCQ to the theoretical expected distortion for both the MDSTCQ and the single description STCQ of [5]. As before, we generate 100,000 triplets and jointly quantize 1000 sinusoidal parameters using a 256 state MDSTCQ with a target entropy of 15 bit/sinusoid per description and a 256 state STCQ with 30 bit/sinusoid into one description and with a packet-loss probability of  $p$ . The distortion is calculated as in the previous experiment. The performance as a function of the packet-loss probability is shown in Fig. 4. From the figure, it can be seen that the theoretical bound of MDSTCQ is better than the theoretical bound of STCQ for a large range of packet-loss probabilities, except when packet-loss probabilities are close to zero. For the practical MDSTCQ, we limit ourselves to one design with  $N = 4$  and compare it to the corresponding MDSTCQ theoretical bound. Typically,  $\Gamma$  is constant and  $\Gamma \approx 3$  for a 256 state [5], but for small  $N$  this is not the case and by simulation  $\Gamma \approx 4$  has been determined. In the figure, a 2.2 dB gap between the practical and the theoretical performance can be seen for low  $p$ . This is most likely the same suboptimality observed in the previous experiment. However, we note that even this constant MDSTCQ design will outperform the single description STCQ for a large range of packet-loss probabilities.

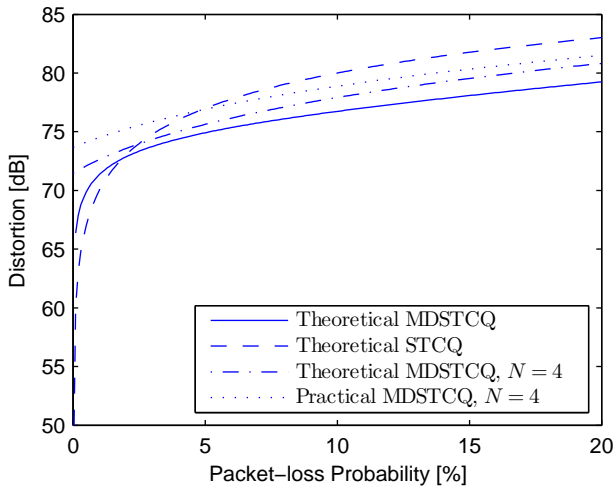


Fig. 4. The theoretical expected distortion of the MDSTCQ (solid) together with the theoretical expected performance of the single description STCQ (dashed). For  $N = 4$ , the theoretical expected distortion of the MDSTCQ (dash-dotted) along with the practical MDSTCQ (dotted).

Audio transmission over unreliable networks results in two distortion contributions, namely quantization distortion and packet-loss distortion. It seems natural to ask whether to use bits on packet-loss concealment or on quantization. To investigate this, the theoretical performance of the STCQ and the MDSTCQ using equal entropy per sinusoid has been plotted in Fig. 5 for two packet-loss probabilities,  $p = 10\%$  and  $p = 20\%$ . From the figure, it can be seen that the STCQ scheme will not be improved beyond approximately 15 bit per sinusoids. This is due to the fact that the quantization distortion is neglectable for  $p = 10\%$  and  $p = 20\%$ . For MDSTCQ, the performance can be improved by adjusting the trade-off between the central and the side distortion. Furthermore, we note that the performance of MDSTCQ also saturates for high entropy indicating that at such rates, the MDSTCQ scheme may benefit from more descriptions.

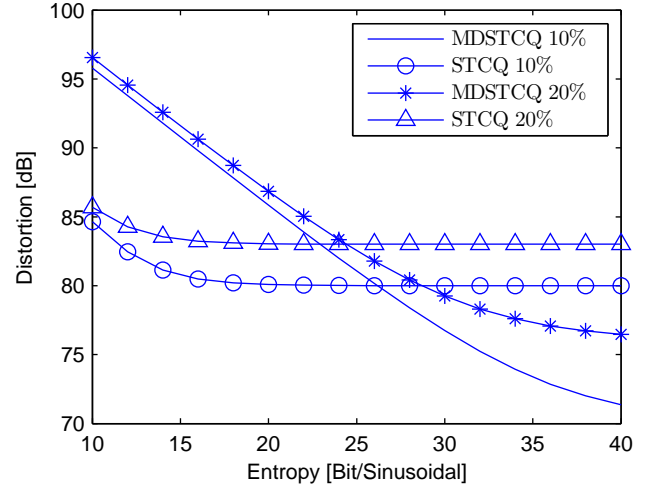


Fig. 5. The theoretical expected distortion of the MDSTCQ at 10% (solid) and 20% (stars) packet-loss probability together with the theoretical expected distortion of the STCQ at 10% (circles) and 20% (triangles) packet-loss probability.

## VI. CONCLUSION

In this correspondence, we have proposed multiple description spherical trellis-coded quantization of sinusoids. The quantizers are suitable for parametric audio coding where the number of sinusoids may vary and for transmission over unreliable networks like the Internet. Under high-resolution assumptions we have derived analytical expressions for the optimal design and the expected perceptual distortion for a given target entropy and packet-loss probability. A desirable feature of the proposed scheme is that the sinusoidal parameters are quantized jointly in a computationally efficient manner and without requiring complex index assignment designing or storage unlike, e.g., [8] and [9]. Experiments have shown significant performance improvements of the proposed scheme as the number of dimensions is increased. Furthermore, a significant performance gain compared to the single description spherical trellis-coded quantization scheme of [5] has been observed for a large range of packet-loss probabilities. We have investigated the importance of the entropy and packet-loss probability for both the proposed scheme and the single description scheme, and experiments have shown that the proposed scheme saturates at lower distortions for high entropies than the single description scheme when packets are lost.

## REFERENCES

- [1] M. G. Christensen and S. H. Jensen, "On perceptual distortion minimization and nonlinear least-squares frequency estimation," *IEEE Trans. Audio, Speech and Lang. Processing*, vol. 14, no. 1, pp. 99–109, January 2006.
- [2] R. Vafin and W. Kleijn, "Entropy-constrained polar quantization and its application to audio coding," *IEEE Trans. Speech Audio Processing*, vol. 13, no. 2, pp. 220–232, March 2005.
- [3] R. Vafin, D. Prakash, and W. Kleijn, "On frequency quantization in sinusoidal audio coding," *IEEE Signal Processing Lett.*, vol. 12, no. 3, pp. 210–213, March 2005.
- [4] P. Korten, J. Jensen, and R. Heusdens, "High-resolution spherical quantization of sinusoidal parameters," *IEEE Trans. Audio, Speech and Lang. Processing*, vol. 15, no. 3, pp. 966–981, March 2007.
- [5] M. H. Larsen, M. G. Christensen, and S. H. Jensen, "Variable dimension trellis-coded quantization of sinusoidal parameters," *IEEE Signal Processing Lett.*, vol. 15, pp. 17–20, 2008.

- [6] V. Goyal, "Multiple description coding: compression meets the network," *IEEE Signal Processing Mag.*, vol. 18, no. 5, pp. 74–93, September 2001.
- [7] R. Aarean, J. Kovacevic, and V. Goyal, "Multiple description perceptual audio coding with correlating transforms," *IEEE Trans. Speech Audio Processing*, vol. 8, no. 2, pp. 140–145, March 2000.
- [8] J. Østergaard, O. Niamut, J. Jensen, and R. Heusdens, "Perceptual audio coding using  $n$ -channel lattice vector quantization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 5, 2006, pp. 197–200.
- [9] G. Schuller, J. Kovacevic, F. Masson, and V. Goyal, "Robust low-delay audio coding using multiple descriptions," *IEEE Trans. Speech Audio Processing*, vol. 13, no. 5, pp. 1014–1024, September 2005.
- [10] J. Østergaard, J. Jensen, and R. Heusdens, "n-channel entropy-constrained multiple-description lattice vector quantization," *IEEE Trans. Inform. Theory*, vol. 52, no. 5, pp. 1956–1973, May 2006.
- [11] J. Kelner, V. Goyal, and J. Kovacevic, "Multiple description lattice vector quantization: variations and extensions," in *Proc. IEEE Data Compression Conference*, 2000, pp. 480–489.
- [12] V. Vaishampayan, A. Calderbank, and J.-C. Batllo, "On reducing granular distortion in multiple description quantization," in *Proc. IEEE Int. Symp. Information Theory*, 1998, p. 98.
- [13] H. Jafarkhani and V. Tarokh, "Multiple description trellis coded quantization," in *Proc. IEEE Int. Conf. Image Processing*, vol. 1, 1998, pp. 669–673.
- [14] —, "Multiple description trellis-coded quantization," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 799–803, June 1999.
- [15] S. van de Par, A. Kohlrausch, R. Heusdens, J. Jensen, and S. H. Jensen, "A perceptual model for sinusoidal audio coding based on spectral integration," in *EURASIP J. on Applied Signal Processing*, vol. 9, 2005, pp. 1292–1304.
- [16] R. Heusdens, J. Jensen, W. B. Kleijn, V. Kot, O. A. Niamut, S. van de Par, N. H. van Schijndel, and R. Vafin, "Bit-rate scalable intra-frame sinusoidal audio coding based on rate-distortion optimisation," *J. Audio Eng. Soc.*, vol. 54, no. 3, pp. 167–188, March 2006.
- [17] M. G. Christensen and S. van de Par, "Efficient parametric coding of transients," *IEEE Trans. Audio, Speech and Lang. Processing*, vol. 14, no. 4, pp. 1340–1351, July 2006.
- [18] M. G. Christensen, A. Jakobsson, S. V. Andersen, and S. H. Jensen, "Amplitude modulated sinusoidal signal decomposition for audio," *IEEE Signal Processing Lett.*, vol. 13, no. 7, pp. 389–392, July 2006.
- [19] C. Tian and S. Hemami, "A new class of multiple description scalar quantizer and its application to image coding," *IEEE Signal Processing Lett.*, vol. 12, no. 4, pp. 329–332, April 2005.
- [20] V. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. Inform. Theory*, vol. 39, no. 3, pp. 821–834, May 1993.
- [21] T. Fischer and M. Wang, "Entropy-constrained trellis-coded quantization," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 415–426, March 1992.
- [22] L. Ozarow, "On a source-coding problem with two channels and three receivers," *Bell Syst. Tech. J.*, vol. 59, no. 10, pp. 1909–1921, December 1980.